feature article

Four Steps to Simplify Multimeasure Performance Evaluations Using the Balanced Scorecard

Thomas L. Albright, Christopher M. Burgess, Aleecia R. Hibbets, and Michael L. Roberts

Processing complex, unstructured information can be difficult. Since balanced scorecards can include as many as 24 or more measures, this article presents a methodology to help managers manage the complexity. The four-step process explained here can lead to accurate and consistent evaluations using both quantitative and qualitative performance measures.

Multiple-item performance evaluation systems, such as the balanced scorecard (BSC), have been advanced as useful tools for performance evaluation within organizations. They commonly are used for evaluating performance of strategic business units and also individuals.

Effective scorecards include metrics selected by top management to represent strategic objectives of the organization; scorecards often include as many as 24 performance measures. These measures should focus on activities and reflect how each position in the organization can impact the organization's strategic objectives.

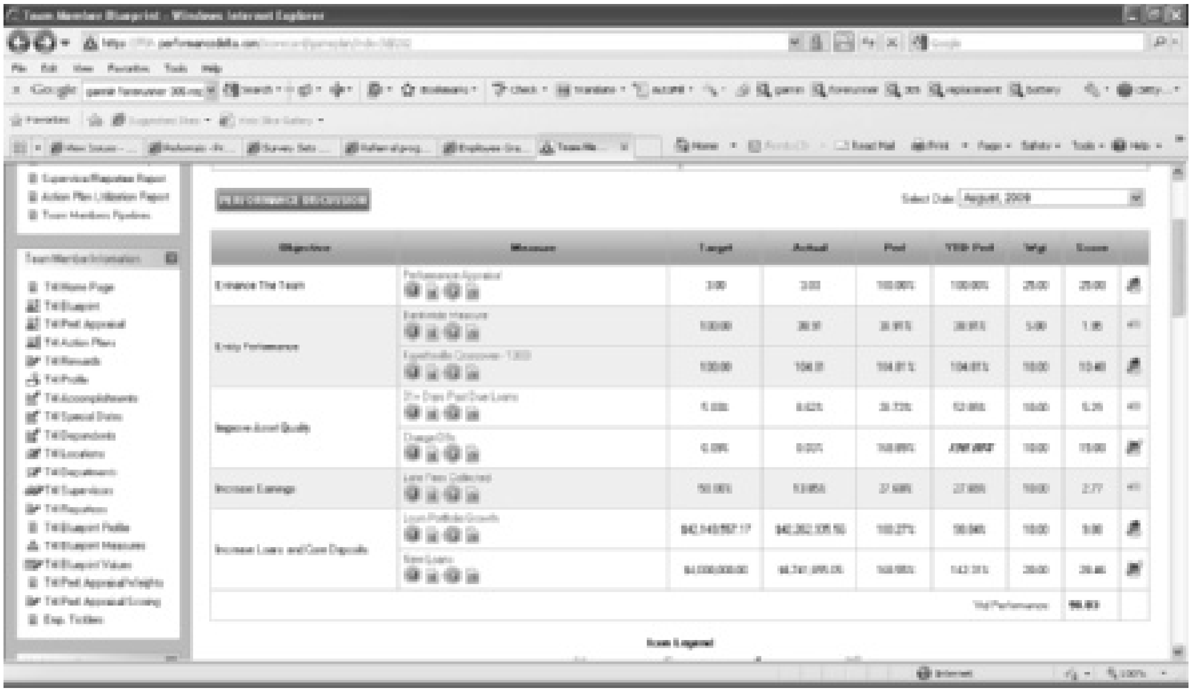
For example, a teller and a loan officer in a banking organization have different impacts on the strategic objectives because of their diverse responsibilities and spheres of influence. Tellers may impact return on assets (ROA) by managing cash over/short each day or by actively pursuing fee collection consistent with company policies. Loan officers, by contrast, can impact ROA by improving the efficiency and effectiveness of the loan process, growing their portfolio of loans outstanding, maintaining or increasing the yield on their portfolios, reducing charged-off and nonperforming loans, or generating low-cost deposit balances from their loan customer base.

As shown in Exhibit 1, a balanced scorecard typically identifies selected performance metrics, plus the subordinate's expected (or target) performance on those measures, and actual performance during the period. The loan officer's scorecard shown in Exhibit 1 contains eight performance metrics associated with five strategically important objectives. Simultaneously evaluating multiple performance measures to arrive at an overall evaluation can be difficult. Managers have inherent limitations on the amount of information that can be simultaneously processed. This situation creates a complex, unstructured decision task. To cope with this difficult task, research has demonstrated that evaluators may ignore some performance information in determining a subordinate's overall evaluation (Lipe & Salterio, 2000). This is problematic, since ignoring some performance metrics undermines the purpose and intent of the

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/jcaf.20613

Exhibit 1

Scorecard Illustrating Objectives, Measures, Target, Actual, Weights, and Score



multimeasure performance evaluation system.

We propose a method for dividing the complex, unstructured overall performance evaluation into a series of individual performance evaluations that are then mathematically combined to derive an overall performance score. This mechanically-derived score can be adjusted, if necessary, to reflect the evaluator's judgment and to arrive at the overall performance evaluation. The goal is to simplify the evaluation process yet improve its effectiveness.

When developing a balanced scorecard, companies typically start with strategic objectives, then define how each position in the organization can influence the objectives. Identifying how each position influences the strategic objectives is known as "identifying influence factors." We advise companies to break the overall influence factors into manageable elements, being careful not to lose focus by including metrics that are not aligned with strategic objectives. Thus, the scorecard should contain a number of meaningful "separate" measures. The four-step simplification of the evaluation process is detailed in the sections that follow.

1. SEPARATELY EVALUATE EACH PERFORMANCE MEASURE

Rather than evaluating a subordinate's overall performance, the scorecard allows managers to evaluate critical components separately. The manager can begin the review process by evaluating the subordinate's performance on each metric compared to the target performance and assigning a score for each item. Most scorecards contain a mix of quantitative and qualitative metrics.

Quantitative Metrics

Typically, quantitative metrics require little or no judgment on the part of the manager. These metrics often are automatically generated by the performance measurement system. Examples that we have used in banking implementations include loan production (number of loans generated), number of past-due accounts, documentation exceptions (number of loan documents, or files, with missing support), and average time from loan application to closing. Typically, targets for quantitative metrics are set by management

DOI 10.1002/jcaf

in a top-down approach. Then, goal action plans are developed throughout the organization to achieve the strategic targets established by management. The balanced scorecards reflect individual results (quantitatively and qualitatively) of the goal action plans.

Qualitative Metrics

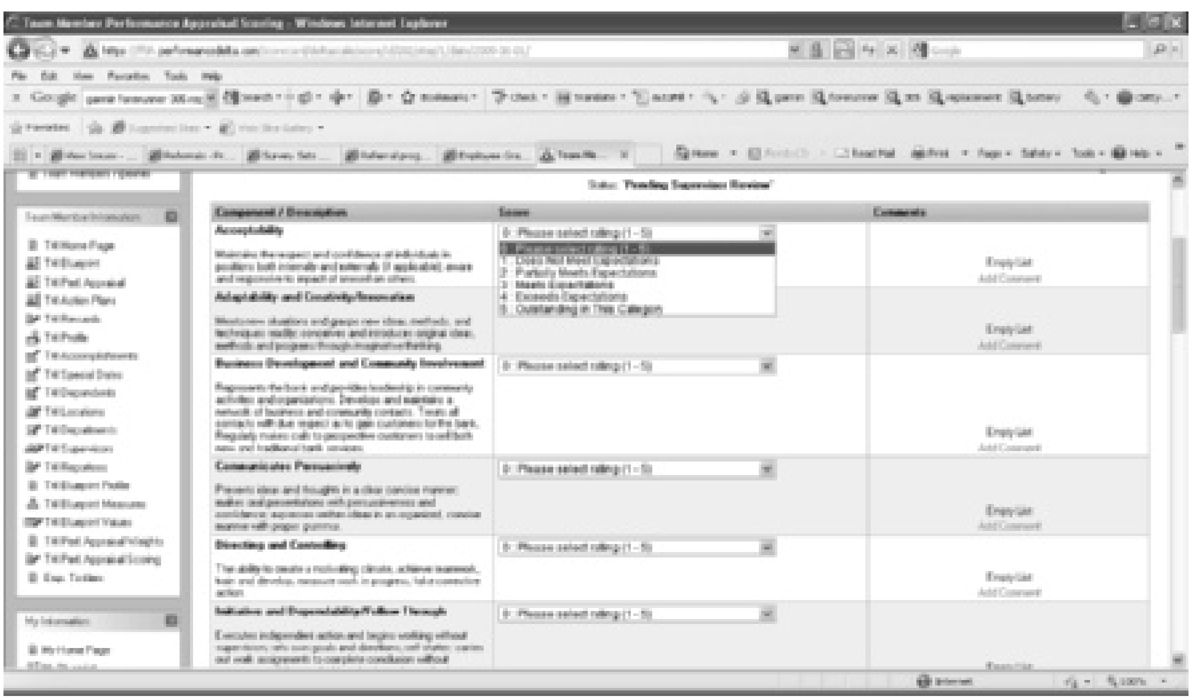
Evaluating performance using qualitative metrics requires judgment, as these metrics often capture behavioral characteristics such as leadership, teamwork, and attitude. Qualitative metrics often are significant drivers of success, yet organizations have difficulty motivating and measuring these behaviors. Many banking institutions use qualitative metrics to set and communicate expectations. For example, smaller banking institutions cannot compete with larger institutions in traditional advertising venues such as TV, radio, and print media. As a result, other methods of establishing and maintaining community relationships are necessary. Thus, many institutions sponsor "after-hours" charity events to connect with the community. Scorecard metrics can encourage desired behavior by measuring and rewarding participation in these events.

Exhibit 2 illustrates the expanded set of qualitative metrics extracted from the scorecard of a loan officer. The qualitative factors in this organization are known, in aggregate, as "the Delta measure" (or "Performance Appraisal" identified in Exhibit 1). As shown below, six strategically important qualitative measures were identified for this position. In developing scorecard measures, it is important to clearly define the characteristics of the performance metric, as shown in the examples below.

- *Acceptability*—Maintains the respect and confidence of individuals in positions both internally and externally.
- *Adaptability and Creativity/Innovation*—Meets new situations and grasps new ideas, methods, and techniques readily.
- *Business Development and Community Involvement*—Represents the bank and provides leadership in community activities and organizations.
- *Communicates Persuasively*—Presents ideas in a clear and concise manner

Exhibit 2

Evaluating Qualitative Factors—The Delta Measure



DOI 10.1002/jcaf

- *Directing and Controlling*—Demonstrates an ability to create a motivating climate and achieve teamwork.
- *Initiative and Dependability/Follow-through*—Executes action and begins working without supervision.

Having identified and defined the strategically important metrics, management must evaluate subordinates using an easily understood, credible, and relevant scale. Scales vary across organizations, depending on the company's objectives. In our experience, a five-point scale commonly is used. Shown below is an example of a five-point scale, together with definitions, that have been successfully used to evaluate performance in a banking environment.

1. Unsatisfactory (consistently and materially below standards).
2. Needs improvement (occasionally, but not consistently, below standards).
3. Successful (meets and occasionally exceeds standards).
4. Excellent (frequently exceeds standards).
5. Distinguished (always exceeds standards).

To help ensure consistency across the organization, managers should be trained to use the scale, and to recognize coaching opportunities based on scorecard results. We have witnessed inconsistencies in scoring across managers. To achieve consistency throughout the organization, some companies require the human resources (HR) department to review all scorecard values. As a result, managers who consistently rate their subordinates high were asked to justify their scores. Often, the HR department discovered an employee was meeting expectations—appropriately evaluated as a 3—yet received a score of 4 or higher. We recommend communicating in advance, in order to anchor the evaluator's expectations and to standardize the evaluation scale. We have often observed managers who are reluctant to assign a 3 out of 5 when an employee achieves 100 percent of target. Yet, a 3 represents "meets expectations." As often found in university settings, grade inflation in a scorecard application can make results difficult to interpret.

Having identified and defined the strategically important metrics, management must evaluate subordinates using an easily understood, credible, and relevant scale.

2. HAVE PREASSIGNED WEIGHTS ATTACHED TO EACH PERFORMANCE MEASURE

As part of designing the performance evaluation system, individual measures are selected and target performance on each measure is determined. In addition, a weight should be assigned to each performance measure. Weights communicate the relative importance of the individual measures and can help the subordinate understand how to allocate his or her effort during the performance period. Using weights also can help bring equity to the performance measurement process. Because individuals have different skills, talents, and abilities, scorecard weights should be specific to the individual, rather than standardized across strategic business units or individuals.

For example, if an employee is talented at developing business, but less so in writing up loan requests, their scorecard may be weighted more heavily for business development measures. Conversely, if an employee is more talented with analysis, and less so with business development, their scorecard could reflect heavier weights for write-ups and loan quality measures. By heavily weighting components (measures), we have seen significant behavioral changes take place. As a result, organizations have benefitted from meeting growth targets.

Additionally, different weights are used when business units within the organization have different objectives, or strategic goals. These goals can be a function of the business unit's life cycle—startup, growth, expansion, or maturity. For example, banking centers located in an economically expanding area would have goals defined by growth in earnings, number of loans, and number of accounts. Alternatively, banking centers located in areas suffering from population declines may emphasize portfolio management rather than growth. By weighting the measures for the business unit, the scorecard not only clearly communicates the role of the business unit but brings fairness to the process. Keep in mind that total weights should not exceed 100 percent for each employee's (or entity's) scorecard. Doing so helps to maintain consistency throughout the development process.

DOI 10.1002/jcaf

3. MATHEMATICALLY DETERMINE AN OVERALL SCORE

The manager can now multiply the individual evaluation on each performance measure by the preassigned weight to determine a weighted score on each item. These individual weighted scores can then be summed, resulting in an overall performance that incorporates performance measures and priorities (weighting) deemed important by senior management. The scorecard process, however, should not end at this point. Management must consider the overall implications and fairness of the mechanically-derived score.

4. ALLOW THE MANAGER TO DETERMINE OVERALL EVALUATION YET REQUIRE JUSTIFICATIONS OF CHANGES TO SCORECARD VALUES

Although the performance evaluation process could end with the mechanically-derived score, it is generally advantageous to allow the supervising manager discretion in determining the overall outcome of the evaluation. The numerical performance measures may reflect results beyond the subordinate's control, of which the supervisor may be aware, or omit information considered important by the supervising manager. To ensure equity, management should have the flexibility to override or make adjustments to the numerical values reported on the scorecard.

Allowing adjustments of the mechanically-derived score can offer advantages and disadvantages.

- *Advantages.* If the external environment changes, unadjusted scorecard values may not permit meaningful performance assessments. For example, in 2009 many banks made the strategic decision to reduce lending activity because of the turmoil in financial markets and the overall economy. Management should be allowed to reduce loan targets that were set prior to the change in strategy. Target values shown on the scorecard may be altered to reflect the new competitive reality.
- *Disadvantages.* Potential abuse can occur in the form of favoritism, or evaluations may be overstated because managers don't want to deliver bad news.

Although the performance evaluation process could end with the mechanically-derived score, it is generally advantageous to allow the supervising manager discretion in determining the overall outcome of the evaluation.

From our experience, we believe it is important to provide flexibility by allowing changes in scorecard values. Since companies do not operate in a vacuum, sometimes circumstances and priorities change. However, if scorecard metrics are changed, it is vital to document the reason. It could also be unfair to other employees if changes are *not* made. For example, if a loan officer inherits a portfolio from another officer (and by doing so greatly exceeds growth targets), to reward this loan officer with a large bonus would be unfair to other employees.

Requiring supervisors to justify modifications of the scorecard is a control against the undesired consequences of subjectivity (Ashton, 1990; Libby, Salterio, & Webb, 2004), while still allowing the flexibility of deviations viewed as necessary by the evaluator. Roberts, Albright, and Hibbets (2004) demonstrated that when weights are preassigned, final evaluations follow the mechanical aggregation of individual performance evaluations and weights very closely. The evaluator has the freedom to modify scorecard weights or target values but is also accountable for the reasoning in doing so. In our experience, organizations require that management obtain permission before altering scorecard weights or target values.

OTHER CONSIDERATIONS

When managers have prepared the mechanically-derived score and made any necessary adjustments, they should consider using the scorecard to direct future activities. The scorecard enhances performance evaluation by allowing managers to walk through the overall assessment and discuss how performance is tied to bonus, incentive, or ranking systems. In addition, the scorecard allows management to document the overall performance process including subsequent discussions. A key element is engaging the employee to an action plan. Managers and subordinates can agree on a target date to accomplish the items in the action plan, the resources

DOI 10.1002/jcaf

required to achieve the targets, potential obstacles, and steps necessary to meet and overcome the obstacles.

SUMMARY AND CONCLUSIONS

Because managers have difficulty processing complex, unstructured information, we believe that a methodology can help simplify the process. Based on our experience with scorecard implementations, we present a four-step process resulting in accurate and consistent evaluations using both qualitative and quantitative metrics:

(1) Evaluate performance on each measure separately.
(2) Use preassigned weights for each performance measure.
(3) Mathematically determine an overall score based on the objective performance measures.
(4) Allow managers some discretion in determining overall performance but require them to justify changes in scorecard metrics.

It has long been recognized that as information processors, human beings have limited capacity for simultaneously processing multiple pieces of information—the famous "seven plus or minus two" limit identified by Miller (1956). Since balanced scorecards often include as many as 24 or more measures (Kaplan & Norton, 1996), a solution to the practical limit of human information-processing ability must be designed into performance-evaluation processes.

Prior research (Roberts et al., 2004) and experience in implementing BSC systems have shown an effective method for overcoming this limitation is to disaggregate the process to evaluate performance on a single measure at a time—breaking the task down into manageable bits of information—and then combining the separate evaluations using preassigned weights. When this method is combined with allowing managers the flexibility to adjust performance scores to achieve greater fairness and requiring managers to justify departures from mathematically-derived scores, all parties involved in the process benefit.

REFERENCES

Ashton, R. H. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback and justification. Journal of Accounting Research, 28 (Supplement), 148–180.

Kaplan, R. S., & Norton, D. P. (1996). The balanced scorecard: Translating strategy into action. Boston, MA: Harvard Business School Press.

Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. The Accounting Review, 79, 1075–1094.

Lipe, M., & Salterio, S. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. The Accounting Review 75, 283–298.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97.

Roberts, M. L., Albright, T. A., & Hibbets, A. R. (2004). Debiasing the balanced scorecard. Behavioral Research in Accounting, 16, 75–88.

Thomas L. Albright is an accounting professor at the University of Alabama in Tuscaloosa. **Christopher M. Burgess** is president and CEO of PerformanceDelta, LLC, in Ridgeland, Mississippi. **Aleecia R. Hibbets** is an accounting professor at the University of Louisiana Monroe. **Michael L. Roberts** is an accounting professor at the University of Colorado at Denver.

DOI 10.1002/jcaf